M_ss_ng D_ta

Janet Wittes
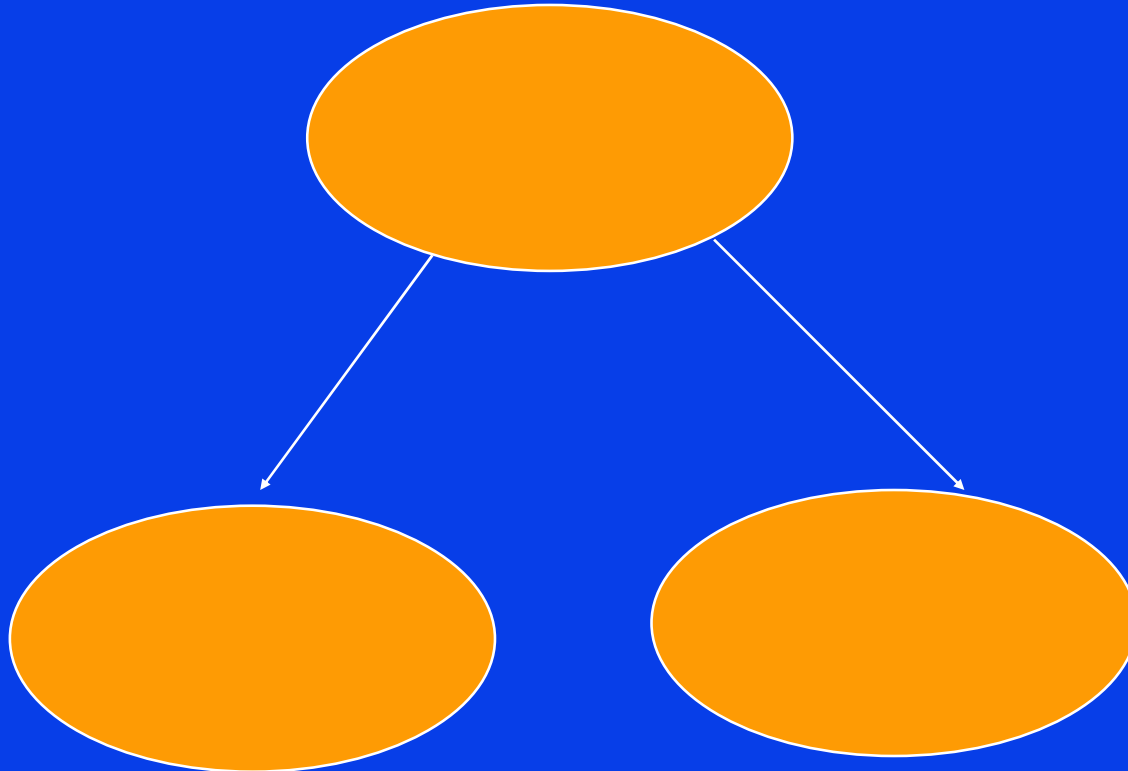**Statistics Collaborative**

_ _ **ssing _at_**
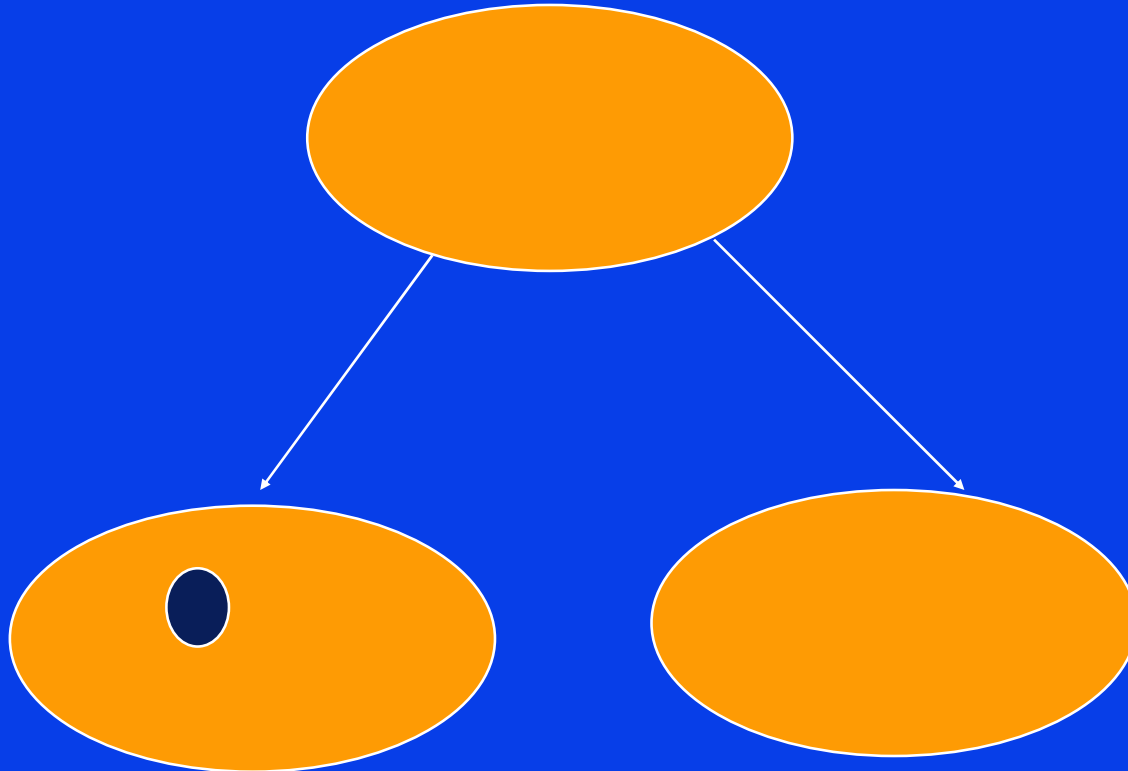
Janet Wittes
Statistics Collaborative

The Caterpillar

- Calling something the "intent-to-treat" population doesn't mean the analysis is "intent-to-treat"

- "one randomized always analyzed"

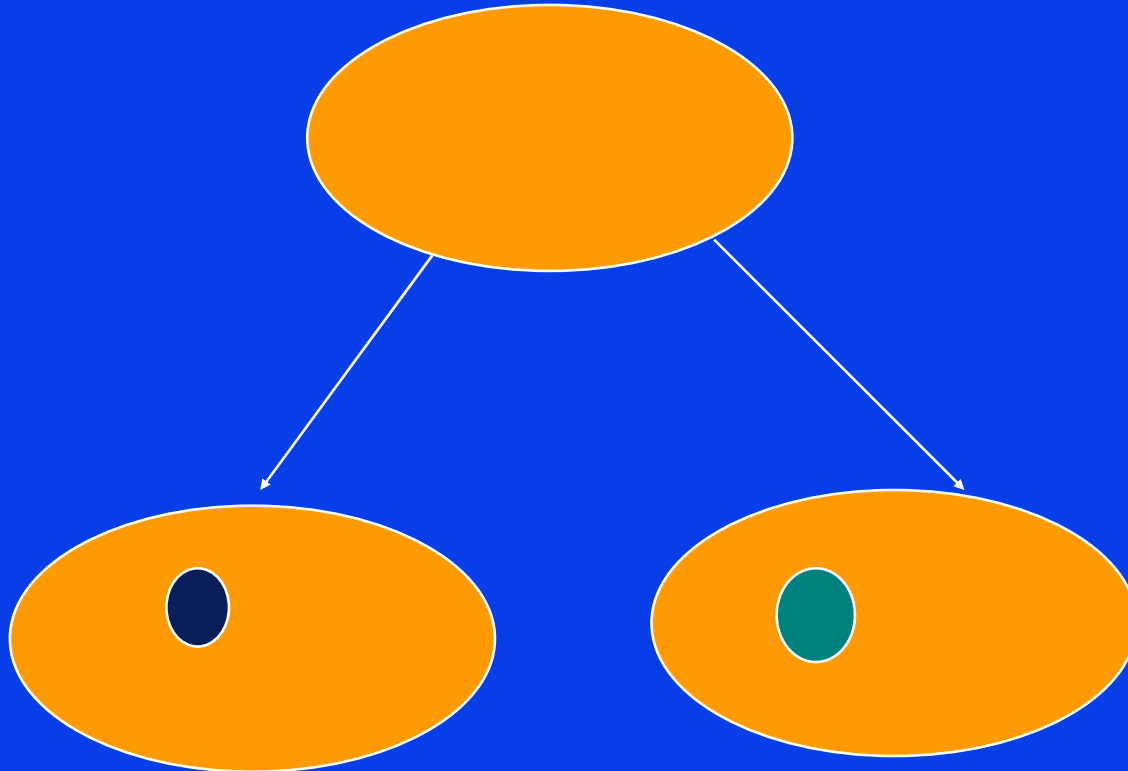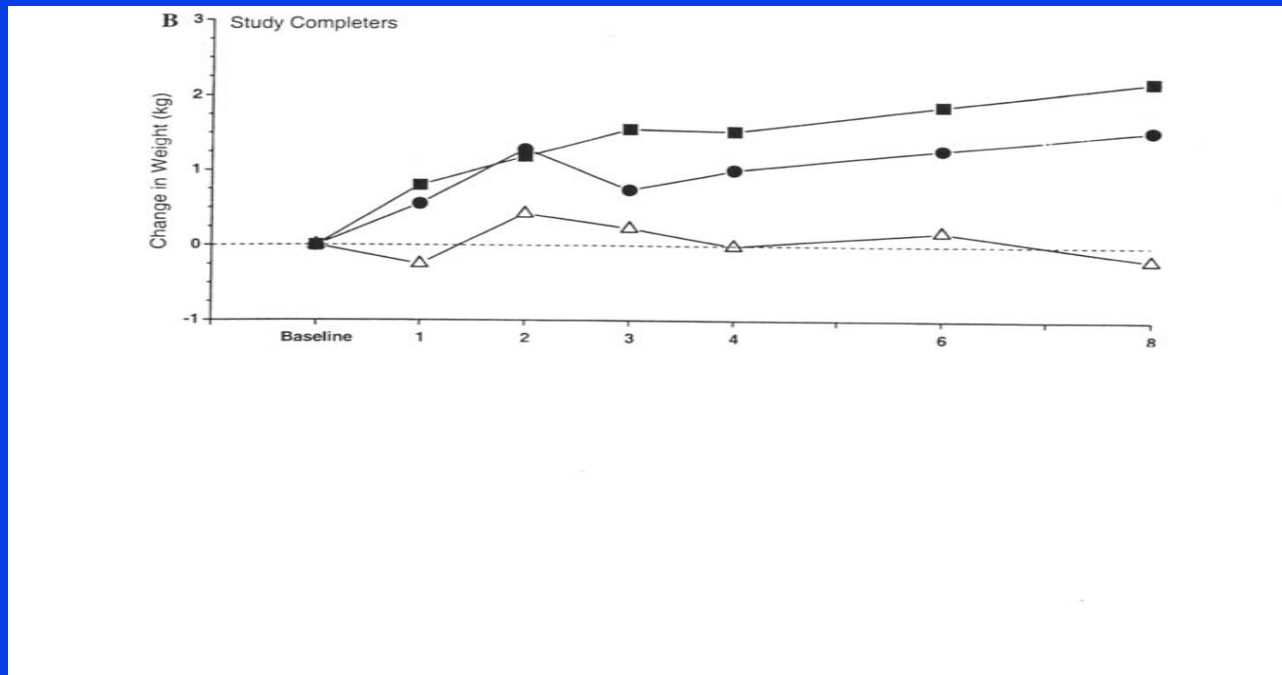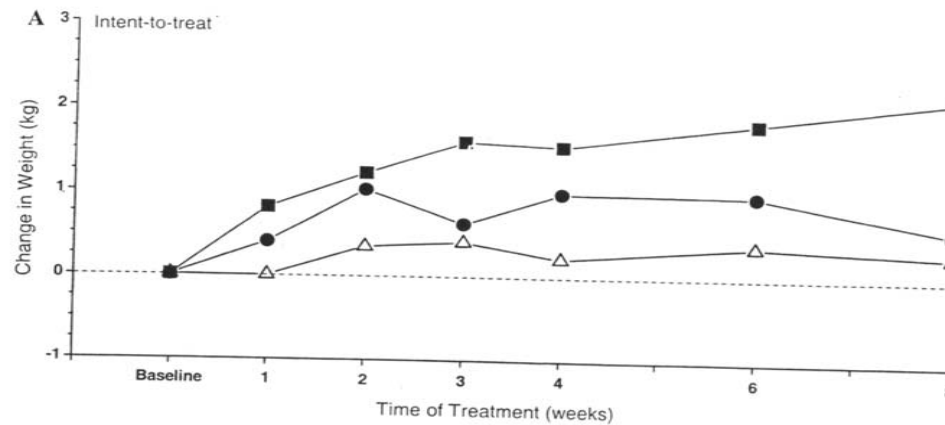- but it depends on the analysis

Selection

Selection

Selection
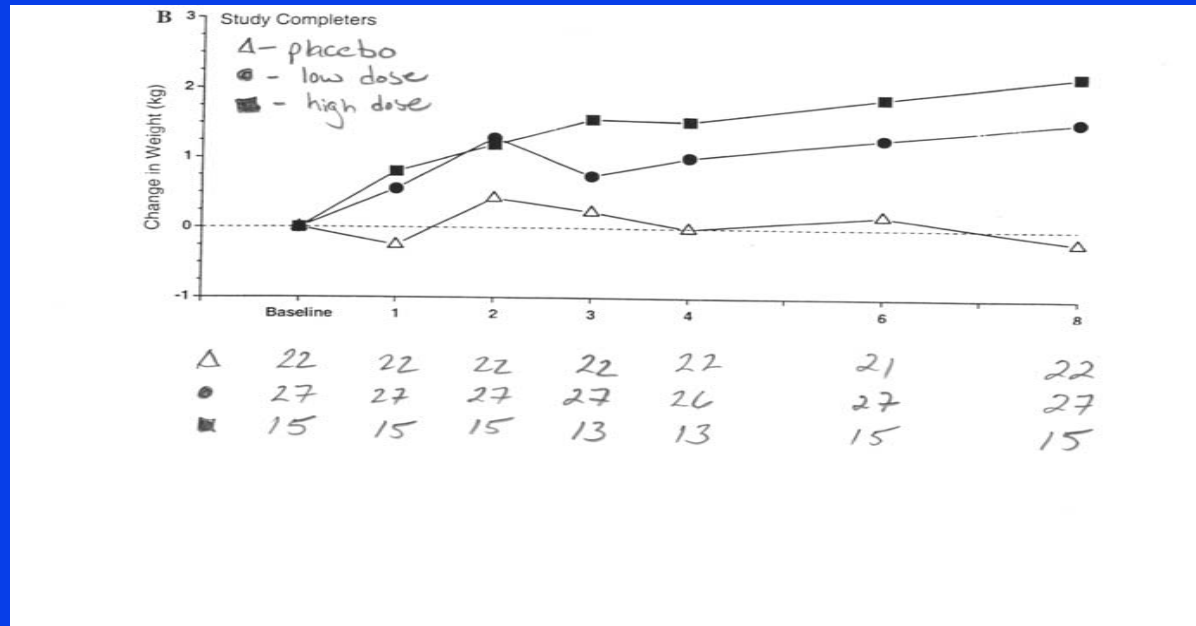
Study Completers

♦=placebo ■ =low ● =high

"Intent-to-treat"

♦=placebo ■ =low ● =high

Completers

♦=placebo ■ =low ● =high

"Intent-to-treat"
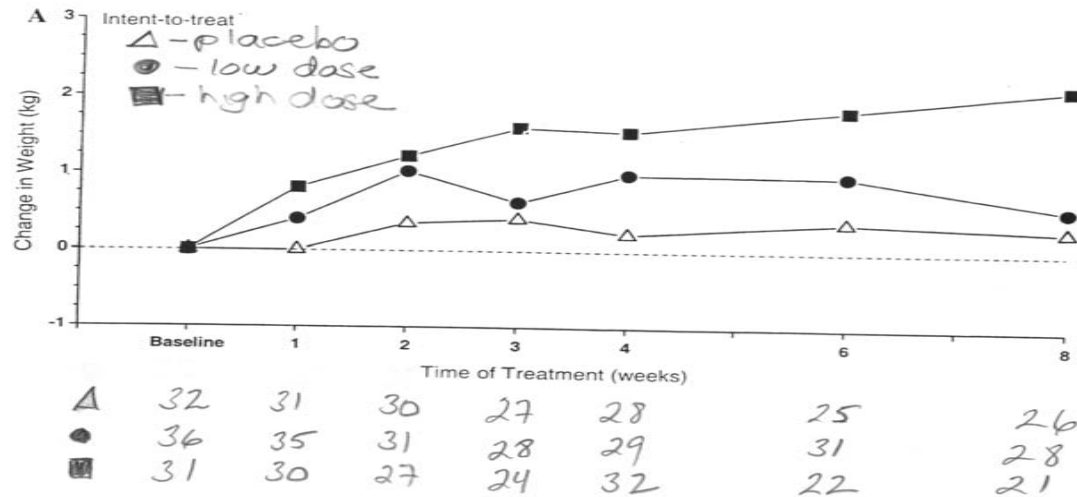
♦=placebo ■ =low ● =high

Daily Mean Number of Urinary Incontinence Episodes

	Placebo	Behavior	Drug
Baseline	5.0 ± 3.2	4.8 ± 2.9	4.7 ± 2.9
Proportion missing final value			
	10%	12%	18%
Change to end of study			
LOCF	-2.1	-2.9	-3.2
Observed	-2.6	-2.8	-3.0

(n=100/group)

Data collected

Completion of study/Withdrawal from study

Is the subject being withdrawn before the planned end of study? .... N☐ No   Y☐ Yes

If NO, please sign and date below and fill out only section "Completion of study".
If YES, please fill out the whole page.

Withdrawal from study

Date of withdrawal: |___|/|___|/|2,0,0,__|
                 (day) (month) (year)

Reason for premature withdrawal from study *(check only one)*
A☐ the subject no longer meets the criteria to remain in the study
B☐ new adverse event or worsening of an existing adverse event
    Please supply an "Adverse event report" form as appropriate.
C☐ lack of efficacy
D☐ poor compliance with treatment
E☐ the subject did not wish to continue in the study
F☐ the subject is lost to follow-up
G☐ administrative reasons
H☐ protocol violation
I☐ the subject died
    Please fax a "Serious adverse event report" and complete "In case of death" form.
L☐ at the discretion of the investigator
J☐ other reason *(please specify)* _____

Choices

- Don't have any missing values

- Use what you have

- Redefine your endpoint

- Use slope -"random regression"

- Impute

  o If so, how?

Avoid missing values

- **Important to get follow-up measures**

  o **Cessation of program not excuse for failing to measure last observation**

Use what you have

Does not respect the randomization

Redefine your endpoint

- **Ventilator use in acute lung injury**

  o **Number of days ALIVE and not on ventilator**

- **Alcoholism**

  o **Number of days of known abstinence**
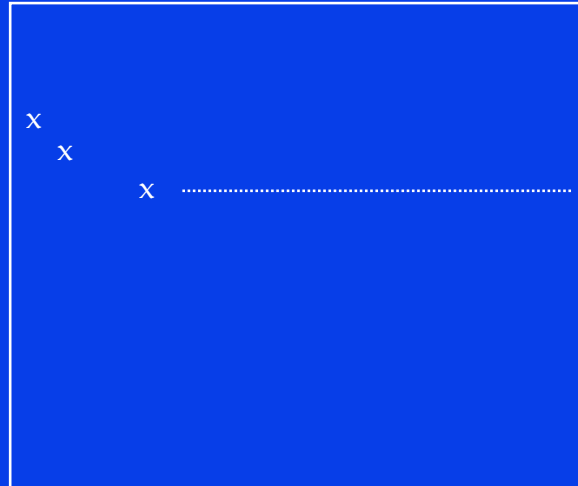
  o **Missing data = heavy drinking**

Redefine…

- **AIDS**

  o **Success=Known increase in weight of at least 1 kg**

- **Incontinence**

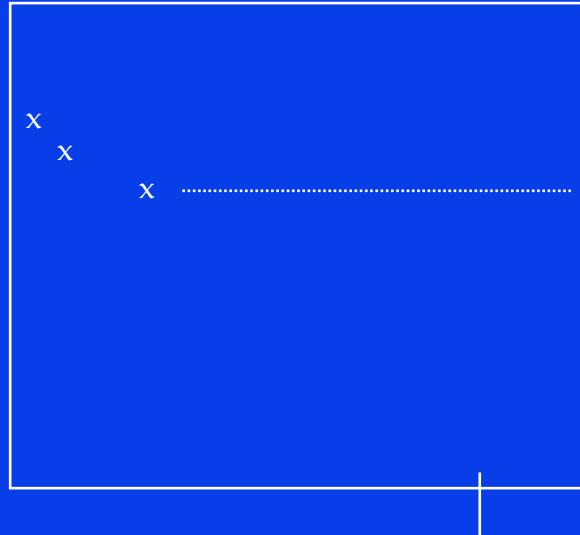 o **Success= Known number of episodes less than 3 times per day**

Impute

- **Idea: assign number to the last value**

- **Choices**

  - o   **LOCF**

  - o   **Windows**

  - o   **Worst case**

  - o   **Worst reasonable case**

  - o   **Multiple imputation**

LOCF

Windows

x
   x
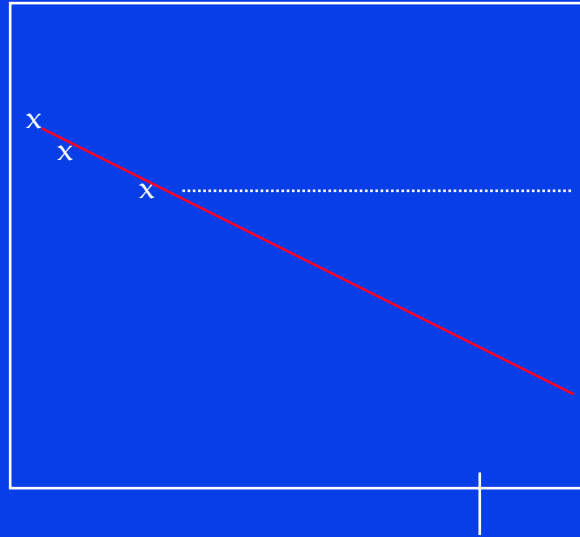   x   ....
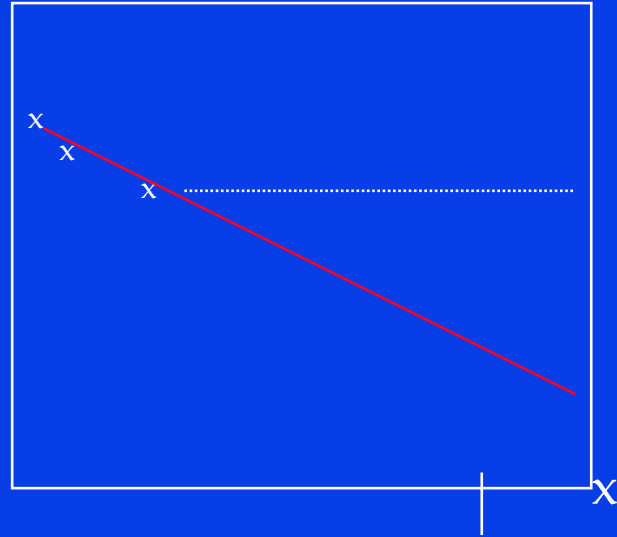
Slope

- **Assume that slope extends beyond last measure**

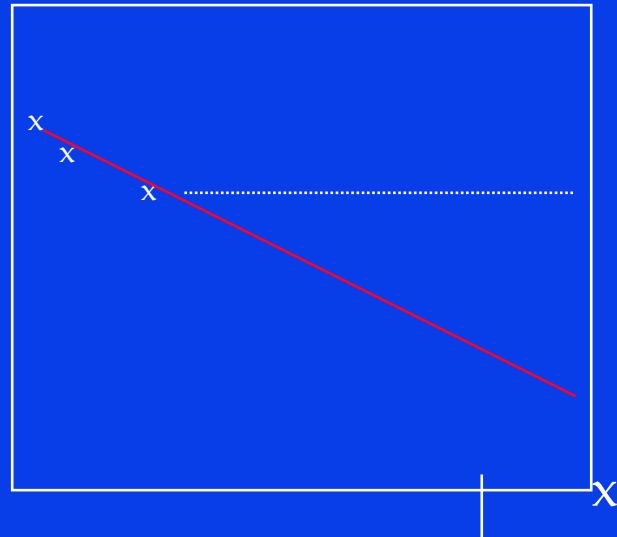 o **Even after death?**

Slope

Worst case

Worst reasonable case

What's wrong with LOCF?

- Too little variability

- Too many degrees of freedom

- Too little variability

Multiple imputation

- Model relationship of baseline variables with outcome

- Simulate outcome many times

- Calculate statistic

- Average

MI vs. LOCF

- **Fixes variability**

- **Fixes df**

- **BUT highly model dependent**

- **And can sometimes give nonsense**

What does it mean to impute?

- **Reason for missing**

  o   **Moved**

  o   **Died**

  o   **Adverse event**

  o   **Quit**

Classify your missing

- Completely at random

- At random

- Related to treatment

"Sensitivity" Analysis

- Do the conclusions vary depending on the method of analysis you use?

AIDS p-values vs. placebo

	Low dose	High dose
Completers	0.006	0.04
Still on original	0.0007	0.10
LOCF	0.012	0.40
Dawson/Lagakos	0.012	0.53
WRC	0.12	0.78
Multiple impute	0.045	0.31

Incontinence example

- **How sure are we that**

  o **The drug works?**

  o **The behavioral intervention works?**

Conclusion

- Think of how much missing data you will have

- Design study to minimize missing data

- In analysis, check how robust your analyses are